UNITED STATES PATENT APPLICATION

FOR

Query Expansion and Weighting Based on Results of Automatic Speech Recognition

INVENTORS:

Benoit Dumoulin

Prepared by:

Blakely, Sokoloff, Taylor & Zafman LLP
12400 Wilshire Boulevard
Seventh Floor
Los Angeles, California 90025
(408) 720-8300

Attorney's Docket No. 3932P022

Query Expansion and Weighting Based on Results of Automatic Speech Recognition

FIELD OF THE INVENTION

The present invention pertains to speech-responsive call routing and information retrieval systems. More particularly, the present invention relates to a method and apparatus for using output of an automatic speech recognizer to improve a query in a call routing or information retrieval system.

BACKGROUND OF THE INVENTION

Call routing systems and information retrieval systems are technologies which help users to identify and select one or more items from among a number of similar items. Call routing systems are commonly used by businesses which handle a large volume of incoming telephone calls. A conventional call routing system uses audio prompts to present a telephone caller with a choice of several selectable options (e.g., topics, people, or departments in an organization). The system then receives a request input by the caller as, for example, dual-tone multiple frequency (DTMF) tones from the caller's telephone handset, associates the caller's request with one of the options, and then routes the call according to the selected option. In a more "open-ended" call routing system, the caller may simply specify a person or other destination and is not limited to a specified set of options.

Information retrieval systems are commonly used on the World Wide Web, among other applications, to assist users in locating Web pages and other hypermedia content. In a conventional information retrieval system, a software-based search engine receives a text-based query input by a user at a computer, uses the query to search a database for documents which satisfy the query, and returns a

list of relevant documents to the user. The user may then select one of the documents in the list to access that document.

Call routing and information retrieval systems can be enhanced by adding automatic speech recognition (ASR) to their capabilities. By using ASR, a user can simply speak his or her request or selection. A natural language speech recognition engine automatically recognizes the user's spoken request and outputs a text-based query, which is then used in essentially the same manner as in a more conventional call routing or information retrieval system. Among other advantages, adding ASR capability to call routing and information retrieval technologies saves time and provides convenience for the user. One problem with using ASR to augment these technologies, however, is the potential for introducing additional error from the ASR process, thus degrading system performance. In a conventional system (i.e., one which does not use ASR), the query is typically input in the form of text, DTMF tones, or some other format which is not particularly prone to error or ambiguity. In contrast, a spoken query may contain both grammatical and syntactical errors (e.g., skipped words, inversions, hesitations). Also, even the best natural language speech recognizers produce recognition errors. Consequently, an ASR-augmented call routing or information retrieval system is susceptible to recognition errors being propagated into the query, reducing the effectiveness of the resulting call routing or information retrieval operation.

SUMMARY OF THE INVENTION

The present invention provides a method and apparatus for identifying one or more items from amongst multiple items in response to a spoken utterance. In an embodiment of the method, an automatic speech recognizer is used to recognize the utterance, including generating multiple hypotheses for the utterance. A query element is then generated based on the utterance, for use in identifying one or more items from amongst the multiple items. The query element includes values representing two or more hypotheses of the multiple hypotheses.

Other features of the present invention will be apparent from the accompanying drawings and from the detailed description which follows.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

Figure 1 is a high-level block diagram of a call routing or information retrieval system which employs ASR capability;

Figure 2 is a flow diagram illustrating a training process associated with the system of Figure 1;

Figure 3 is a flow diagram showing a run-time call routing or information retrieval process associated with the system of Figure 1; and

Figure 4 is a flow diagram showing a process for generating a query based on an n-best list output by a speech recognizer.

DETAILED DESCRIPTION

A method and apparatus for using the output of an automatic speech recognizer to improve a query in a call routing or information retrieval system are described. Note that in this description, references to "one embodiment" or "an embodiment" mean that the feature being referred to is included in at least one embodiment of the present invention. Further, separate references to "one embodiment" in this description do not necessarily refer to the same embodiment; however, neither are such embodiments mutually exclusive, unless so stated and except as will be readily apparent to those skilled in the art. Thus, the present invention can include any variety of combinations and/or integrations of the embodiments described herein.

The technique described below can be used to improve call routing and information retrieval systems which employ automatic speech recognition (ASR). Briefly, the information retrieval or call routing process is made more accurate by forming an expanded query from all of the hypotheses in the n-best list generated by the speech recognizer, and by weighting the query with the confidence scores generated by the recognizer. More specifically, and as described further below, the ASR process is used to recognize a user's utterance, representing a query. The ASR process includes generating an n-best list of hypotheses for the utterance. A query element is generated, containing values representing all of the hypotheses from the n-best list. Each value in the query element is then weighted by hypothesis confidence, word confidence, or both, as determined by the ASR process. The query element is then applied to the searchable items to identify one or more items which satisfy the query.

Figure 1 is a high-level block diagram of a system which may be used to for call routing, information retrieval, or both, in accordance with the present invention. The system includes an audio front end (AFE) 1, an ASR subsystem 2, an information retrieval (IR)/call routing engine 3, and a database 4. The database 4 stores the set of destinations, documents, or other types of data (hereinafter simply "destinations") that can be searched on and potentially selected by a user. The destinations contained in database 4 may be essentially any type of information, such as text or audio or both.

The audio front end 1 receives speech representing a query from the user via any suitable audio interface (e.g., telephony or local microphone), digitizes and endpoints the speech, and outputs the endpointed speech to the ASR subsystem 2. The audio front end 1 is composed of conventional components designed for performing these operations and may be standard, off-the-shelf hardware and/or software.

The ASR subsystem 2 performs natural language speech recognition on the endpointed speech of the user, and outputs a text-based query vector to the information retrieval/call routing engine 3. The ASR subsystem 2 contains conventional ASR components, including a natural language speech recognition engine, language models, acoustic models, dictionary models, user preferences, etc. The information retrieval/call routing engine 3 receives the query vector and, in a conventional manner, accesses the database 4 to generate a list of one or more results which satisfy the query. Accordingly, the information retrieval/call routing engine 3 includes conventional components for performing vector-based call routing and or information retrieval.

It will be recognized that certain components shown in Figure 1, particularly the ASR subsystem 2 and the information retrieval/call routing engine 3, may be implemented at least partially in software. Such software may be executed on one or more conventional computing platforms, such as personal computers (PCs), workstations, or even hand-held computing devices such as personal digital assistants (PDAs) or cellular telephones. These components may also be distributed across one or more networks, such as the Internet, local area network (LANs), wide area networks (WANs), or any combination thereof. Likewise, these components may be implemented at least partially in specially-designed hardwired circuitry, such as application specific integrated circuits, programmable logic devices (PLDs), or the like. Thus, the present invention is not limited to any particular combination of hardware and/or software.

Operation of the system may be categorized into two phases: training and run-time. The system employs a standard vector-based approach, which is augmented according to the present invention. One example of a vector-based call routing approach which may be used for purposes of this invention is described in J. Chu-Carroll et al. "Vector-Based Natural Language Call Routing", Computational Linguistics, vol. 25, no. 3, September 1999, which is incorporated herein by reference.

Prior to run-time, the system of Figure 1 is trained using standard vector-based call routing/information retrieval techniques. More specifically, training may be accomplished by using the technique of latent semantic indexing (LSI). The LSI technique is described in S. Deerwater et al., "Indexing by Latent Semantic Analysis", Journal of the American Society for Information Science, 41(6), pp. 391-407 (1990), which is incorporated herein by reference. The technique includes building an MxN

term-destination matrix containing the frequency of occurrence of terms (word n-grams) for each destination, where M represents the total number of distinct terms (and rows in the matrix) in all of the destinations to be searched and N represents the total number of destinations (and columns in the matrix). Thus, the term-destination matrix contains a set of values, each of which represents the frequency of occurrence of a particular term in a particular destination. The term-destination matrix is then weighted according to a standard weighting technique used in call routing or information retrieval, such as inverse document frequency (IDF). The dimensionality of the matrix is then reduced using a standard technique such as singular value decomposition (SVD).

Figure 2 illustrates the LSI training process according to one embodiment. Initially, at block 201, the term-destination matrix is constructed. At block 202, the term-destination matrix is normalized so that each term vector (row of the matrix) is of unit length (i.e., by dividing each value by the number of values and its row). Next, the matrix is further weighted using IDF at block 203. IDF involves weighting the value for each term inversely to the number of documents in which the term occurs, as is well-known in the art. At block 204, the weighted matrix is reduced in dimensionality using SVD. The SVD process produces two matrices, i.e., an MxN "transformation" matrix and an NxN matrix, the columns of which represent the eigenvectors.

Figure 3 illustrates the run-time process of the system, according to one embodiment. Initially, an utterance (a request) is received from the user at block 301. The utterance is endpointed at block 302, and then recognized by the ASR subsystem 2 at block 303. As noted above, a result of the ASR process is the generation of an n-

best list of hypotheses for the utterance. At block 304, a text-based query vector is formed by the information retrieval/call routing engine 3 based on the recognized utterance, as described further below. At block 305, one or more destinations in database 4 which satisfy the query are identified by the information retrieval/call routing engine 3. Those identified items which meet a predetermined threshold for similarity to the query vector are then indicated and/or provided to the user at block 306. In the case of a call routing system, this may involve simply connecting the user to the destination which most closely matches the query. In an information retrieval system, this may involve outputting to the user a list of "hits", i.e., destinations which most closely match the query.

Figure 4 shows the process of forming a query vector (block 304) in greater detail, according to one embodiment. At block 401, the information retrieval/call routing engine 3 forms a query vector from all of the hypotheses in the n-best list resulting from the speech recognition process. This operation may involve simply concatenating all of the hypotheses in the n-best list and then representing this result in standard vector form. Note that while it may be preferable to use all of the hypotheses in the n-best list to form the query vector, it is not necessary to do so. In other words, the system may use more than one, but not all, of the hypotheses in accordance with the present invention; this would still provide a performance improvement over prior approaches.

Assume now that a user actually uttered the phrase "books about the Internet". Assume further that the ASR subsystem generates a simple three-hypothesis n-best list as follows:

Hypothesis no.	Hypothesis
1	books about the Internet
2	looks about the Internet
3	books of the Internet

Although an actual n-best list would likely be longer than this one, a simple example is chosen here to facilitate explanation. Thus, in accordance with the present invention, the ASR subsystem 2 would generate a query vector representing the concatenation of all three hypotheses, i.e., "books about the Internet looks about the Internet books of the Internet".

Next, at block 402 each value in the query vector is weighted according to its hypothesis confidence, i.e., according to the confidence score, or rank, of the hypothesis to which the value corresponds. For example, the individual vector values which represent the phrase "books about the Internet" are each assigned the highest weight, because that phrase corresponds to the highest-ranked hypothesis in the n-best list, i.e., the hypothesis that the highest confidence level. In contrast, the values representing the phrase "books of the Internet" are assigned the lowest weight, because that phrase corresponds to the lowest-ranked hypothesis, i.e., the hypothesis with the lowest confidence level.

Next, at block 403, optionally, each value in the query vector is further weighted according to its word confidence, i.e., according to a confidence score of the particular word which the value represents. The word confidence score may be, for example, a measure of the number of times the word occurs in the n-best list. For example, the word "Internet" appears in all three of the hypotheses in the above n-best list, and accordingly, would be assigned the highest relative weight in this operation. In contrast, the word "of" occurs only once in the n-best list, and therefore

would be assigned the lowest weight in this operation. Of course, other ways of measuring word confidence are possible, as will be recognized by those skilled in the art.

Any additional standard weighting techniques can then optionally be applied at block 404, such as may be used in other vector-based information retrieval or call routing approaches. Finally, at block 405, a pseudo-destination vector is formed from the query vector by reducing the dimensionality of the query vector from M to N, to allow a similarity comparison with the transformed term-destination matrix.

Referring again to the overall run-time process of Figure 3, in one embodiment block 305 involves comparing the pseudo-destination vector to each of the N eigenvectors to determine which is the closest, according to a standard dot product measure (i.e., cosine score between the vectors). Of course, other methods may alternatively be used to determine similarity between the eigenvectors and the query, such as using the Euclidean distance or the Manhattan distance between the vectors.

Thus, combining the LSI technique with IDF corrects for grammatical and syntactical errors in the query. The use of all (or at least more than one) of the hypotheses in the n-best list in forming the query reduces call routing or information retrieval errors due to speech recognition errors.

Thus, a method and apparatus for using the output of an automatic speech recognizer to improve a query in a call routing or information retrieval system have been described. Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader

spirit and scope of the invention as set forth in the claims. Accordingly, the specification and drawings are to be regarded in an illustrative sense rather than a restrictive sense.